UNIVERSIDAD AUTÓNOMA DE MADRID
ESCUELA POLITÉCNICA SUPERIOR



## Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación

TRABAJO FIN DE GRADO

# DETECCIÓN DE PERSONAS EN ENTORNOS RESIDENCIALES Y HOSPITALARIOS

**Jesús Molina Merchán**

**Tutor: Rafael Martín Nieto**

**Ponente: José María Martínez Sánchez**

**Mayo 2016**

# DETECCIÓN DE PERSONAS EN ENTORNOS RESIDENCIALES Y HOSPITALARIOS

**Jesús Molina Merchán**

**Tutor: Rafael Martín Nieto**

**Ponente: José María Martínez Sánchez**

**Video Processing and Understanding Lab**
**Departamento de Ingeniería Informática**
**Escuela Politécnica Superior**
**Universidad Autónoma de Madrid**
**Mayo 2016**

Trabajo parcialmente financiado por el Ministerio de Economía y Competitividad del Gobierno de España bajo el proyecto TEC2014-53176-R (HAVideo) (2015-2017)

# Resumen y palabras clave.


## Resumen

Existe una amplia demanda en el área de la video-seguridad, sobre todo en la detección de personas, lo que ha provocado un gran aumento del número de investigaciones en este campo.

En el ámbito de la detección de personas mayores, el detector debe de tener en cuenta diferentes posturas como la de estar sentado o en silla de ruedas. También es importante considerar el coste que conlleva realizar estos detectores.

Por todo ello, este trabajo tiene dos objetivos principales. El primero de ellos ha sido elaborar un modelo de persona sentada con el fin de completar un detector en el escenario de una residencia de ancianos. El segundo de ellos se basaba en reducir la cantidad de recursos necesarios, y ahorrar el coste de tener que grabar secuencias para realizar un detector en este escenario, y para ello se han creado tres dataset de imágenes sintéticas de personas en silla de ruedas con el fin de realizar tres modelos diferentes, estudiar qué modelo es el óptimo y por último estudiar su viabilidad comparándolo con el detector de personas en silla de ruedas.



## Palabras clave

Detección de personas, entornos residenciales, entornos hospitalarios, imágenes sintéticas

# Abstract and Keywords


## Abstract

There is a large demand in the area of video-surveillance, especially in detecting people, which has caused a large increase in the number of investigations in this field.

In the field of detection elderly people, the detector must have into account different positions such as sitting or in a wheelchair. Also is important the cost involved in making these detectors.

Therefore, this work has two main objectives. The first has been to develop a model person sitting with the aim of completing a detector on the stage of a nursing home. The second one was based on reducing the amount of resources needed and save the cost of having to record sequences for a detector in this scenario.

To achieve this, three 'synthetic images dataset were created in order to perform three different models, evaluating which model is optimal and finally analizing its feasibility by comparing it with the people detector in wheelchairs.



## Keywords

People detection, nursing homes, hospital environments, synthetic images

# Agradecimientos.

*Quiero agradecer en primer lugar a mi tutor, Rafa, por la confianza depositada en mí y por toda la ayuda que he recibido. En segundo lugar quería agredecer a Álvaro por estar siempre que lo necesitaba y ayudarme en todo. También dar las gracias a todo el VPU Lab por echarme una mano siempre que se lo he pedido.*

*No puedo mas que estar agradecido a Beatriz, por se la voz que me decía puedes cuando más cuesta arriba parecía la situación. Y gracias a ti, pude.*

*Muchas gracias Belén, por sacarme del paso en tantas y tantas ocasiones, pero sobre todo por tu compañía. También quería dar las gracias a todo mi grupo de la universidad, por todos los buenos momentos.*

*Agradecer a toda mi familia, en especial a mi madre. Gracias a ella que me ayudado todos estos años, y por todo el esfuerzo realizado para poder llegar hasta aquí.*

*Por último, agradecer a Rafa, Óscar, Alberto, Rodri y todos los Sabios por toda una vida de amistad.*

*Gracias a todos.*

Jesús Molina

Mayo 2016.

# Índice general

# Índice de figuras

# Índice de tablas

# Capítulo 1

# Introducción.

## 1.1. Motivación.

En la actualidad, el campo de la vigilancia y la seguridad está generando gran interés, lo que ha provocado un gran aumento de los estudios e investigaciones en video-vigilancia y video-monitorización. En este ámbito, cobra especial importancia la detección de personas, una de las principales áreas de estudio de la visión por ordenador.

Estos detectores varían segun el entorno en el que se pruebe, ya que tienen una gran dependencia a factores como la iluminación, las oclusiones, la postura de la persona, la perspectiva, la distancia a la que se encuentra de la cámara, etc.

En especial este trabajo se centra en la detección de personas en entornos hospitalarios y residenciales. Resulta interesante estudiar más en profundidad diferentes apariencias como puede ser sentado o en silla de ruedas.

La motivación de este proyecto será realizar un modelo de personas sentadas, y evaluar su funcionamiento respecto otros modelos como el modelo de silla de ruedas o el de persona erguida en un entorno determinado, para así estudiar la aportación del modelo creado en la detección. Por otro lado, otro motivo por el cual se elaboró este trabajo fue la de estudiar la viabilidad de realizar un detector de personas en silla de ruedas basado en modelos entrenados con imágenes sintéticas.

## 1.2. Objetivos.

Este trabajo tiene dos objetivos principales:

- Realización de un modelo de persona sentada, y la evaluación del detector de persona levantada, detector de persona en silla de ruedas y el detector de persona sentada creado. También se realizará un detector que combine los tres detectores anteriores, y se comentará la posible mejora otenida frente al detector de persona levantada y de persona en silla de ruedas para un entorno residencial.
- Creación de un detector de personas en silla de ruedas basado en imágenes sintéticas. Para ello, inicialmente se crearán imágenes a partir de la combinación de personas erguidas y de imágenes de silla de ruedas de tres formas diferentes: superponiendo el tronco superior y la parte inferior de las piernas en la imagen de la silla de ruedas, con el método anterior y emborronando las zonas de transición de las dos imágenes combinadas, y aplicando una máscara de tal manera que sólo seleccione el cuerpo de la persona y elimine la información de fondo. Una vez obtenidos estos datasets, se observarán los resultados con el fin de comprobar que la diferencia entre el rendimiento los detectores creados y el rendimiento del detector con personas en silla de ruedas con imágenes reales.

## 1.3. Estructura de la memoria.

La memoria del proyecto se divide en los siguientes capítulos:

- Capítulo 1. Introducción: motivación, objetivos y estructura de la memoria.
- Capítulo 2. Estado del arte: sistema de detección de personas y sistema de detección de personas en silla de ruedas.
- Capítulo 3. Desarrollo: proceso seguido para la creación del sistema de detección de personas sentadas y para el sistema de detección de personas en silla de ruedas con imágenes sintéticas.
- Capítulo 4. Experimentos realizados y resultados obtenidos en ellos.
- Capítulo 5. Conclusiones y trabajo futuro.
- Referencias.

# Capítulo 2

# Estado del arte.

Este capítulo proporciona una visión general del trabajo realizado previamente en las áreas relacionadas con el objetivo de este trabajo. Se dividirá en dos secciones, en la primera sección se describirá el funcionamiento de los detectores de personas (sección 2.1), y en la segunda sección se hará lo propio con los detectores de personas en silla de ruedas (sección 2.2).

## 2.1. Detección de personas.

La detección de personas se ha convertido en uno de las áreas de investigación que mayor interés suscita en el ámbito de procesamiento de imagen y vídeo. Se han desarrollado varios sistemas de detección diferentes, sin embargo tal y como se explica en [1], la mayoría de detectores de personas tienen una arquitectura común, que consiste en el diseño y el entrenamiento de un modelo de persona, basándose en ciertos parámetros característicos, como puede ser el movimiento, la silueta o la postura. El siguiente paso, se centra en adaptar dicho modelo a todos los posibles candidatos a ser persona de la escena, y por útlimo si dicho candidato se ajusta al modelo, entonces será clasificado como persona, mientras que los que no se ajusten no serán clasificados como tal.

### 2.1.1. Arquitectura de los sistemas de detección de personas

A continuación se desribirán las etapas principales de la arquitectura de un detector de personas:

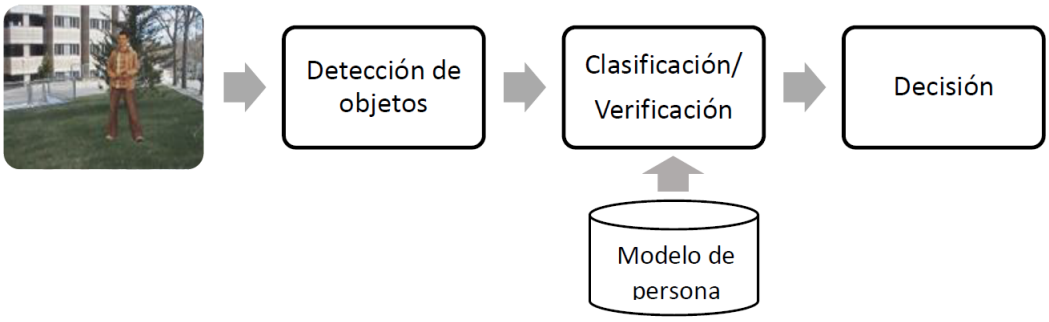



Figura 2.1: Proceso general de detección de personas. Fuente:[2]

- Entrada: hay muchos formatos posibles, sin embargo, para la visión por ordenador la unidad básica de entrada son imágenes o frames.
- Detección de objetos: consiste en generar o extraer los posibles candidatos iniciales a ser persona. Es una tarea crítica para el detector. Se explicará con mayor profundidad en la subsección 2.1.2.
- Modelo de persona: define las características y normas que los objetos deben de cumplir para ser considerados como persona. Junto con la anterior, son las etapas más importantes para la detección de personas. En la subsección 2.1.3 se ahondará más en esta etapa.
- Verificación o clasificación: tiene el mismo funcionamiento que un detector de patrones. Compara modelos entrenados anteriormente y el modelo generado de la secuencia.
- Decisión: según lo obtenido en el paso anterior, toma una decisión de si es o no persona.

### 2.1.2. Detección de objetos

Según [2], hay dos enfoques principales en la detección de objetos, la primera de ellas, la segmentación, se centra en la información de primer plano y segundo plano, y la segunda se basa en la exploración exhaustiva. Ambos enfoques se explicarán a continuación con mayor detalle. Pese a tener distintos enfoques, el resultado final para ambos es la localización y la dimensión de los diferentes objetos detectados en el escenario.

- Segmentación: se emplea para dividir la imagen en regiones separadas, que idealmente corresponden a diferentes objetos del mundo real. De forma más precisa, este proceso trata de asignar una etiqueta a todos los píxeles, de tal forma que píxeles con la misma etiqueta comparten alguna característica visual u otra propiedad, como puede ser el color, el movimiento, la textura, etc. Regiones contiguas deben tener diferencias muy significativas con respecto a la misma característica para ser considerada diferente. El resultado final es, idealmente, localizar y discriminar objetos del primer plano respecto del fondo, como se hace en [3].

- Búsqueda exhaustiva: consiste en un barrido de la imagen entera en búsqueda de similitudes con el modelo de persona elegido, a diferentes escalas y en diferentes posiciones. Con este modelo se obtiene un mapa de confianza muy denso, por lo que para llegar a detecciones individuales se debe buscar máximos locales en el volumen de densidad, y a continuación, aplicar algún tipo de supresión de los no-máximos. Existen dos técnicas para ello, la primera obtiene este volumen de densidad evaluando diferentes ventanas de detección con clasificador, como es el caso de los detectores basados en ventana deslizante, mencionado en [4], mientras que la segunda crea este volumen de densidad de una manera explícita de abajo hacia arriba mediante votaciones probabilísticas emitidas por las características locales equivalentes. Esta técnica es la utilizada por los detectores basados en características, como pueden ser los comentados en [5, 6].

- Segmentación y búsqueda exhaustiva: combina las dos técnicas anteriores, tratando de aprovechar sus respectivas ventajas. Selecciona en una primera vuelta unos candidatos iniciales con el método de segmentación y luego realiza una segunda vuelta mediante búsqueda exhaustiva.

### 2.1.3. Modelos de persona

Como se explicó en la subsección 2.1.1, el proceso de clasificación o verificación utiliza un modelo de persona previamente entrenado, y lo aplica a los candidatos a ser persona de una imagen o una secuencia para determinar una decisión basándose en su similitud. De esta manera, como se expone en [2] contar con un modelo de persona apropiado, es un aspecto crítico en para el proceso de clasificación o verificación. Hay dos fuentes de información para caracterizar el modelo de persona: la apariencia y el movimiento.

- Basados en movimiento: la apariencia humana varía dependiendo de ciertas condiciones como la luz, la vestimenta, el contraste con el entorno, etc. Por esta razón, algunos detectores, como [7], se basan exclusivamente en la información de movimiento. Este algoritmo segmenta el movimiento, realiza un seguimiento de los objetos en el fondo, alinea cada objeto a lo largo del tiempo y finalmente calcula la semejanza entre los objetos y como estos evolucionan en el tiempo.

- Basados en apariencia: esta información es mucho más discriminativa que el movimiento, se clasifican los modelos de apariencia de acuerdo con modelos simplificados de persona, o modelos complejos. Existen modelos simples que definen una persona como una región, por ejemplo modelos holísticos, y otros más complejos que definen una persona como una combinación de múltiples regiones, como pueden ser los modelos basados en partes.

- Basados en apariencia más movimiento, como se explica en [8], en el que se propone un sistema colaborativo.

### 2.1.4. Factores críticos en la detección de personas

Como se ha comentado anteriormente en la subsección 2.1.3, la detección de personas consiste básicamente en diseñar y entrenar un modelo de persona basado en parámetros característicos, y en ajustar dicho modelo a los candidatos en la escena. Por lo tanto, la detección de personas se puede separar en la localización de candidatos iniciales en la escena y su posterior clasificación. Teniendo en cuenta estas premisas, se puede concluir que el proceso global está fuertemente ligado a una serie de propiedades de los objetos y del fondo, y a las propiedades que relacionan ambos elementos. Estas dependencias son las que se han denominado «factores críticos», enfatizando así, en su influencia sobre el resultado final. En la figura 2.2 se muestra una clasificación de algunos de estos factores.

## 2.2. Detector de usuarios en silla de ruedas

Hay algunos trabajos en el estado del arte que tratan de abordar esta técnica. Sendos trabajos se pueden dividir en dos grupos principales. El primero de ellos se centra en detectar elipses que se corresponden con las ruedas de la silla, aunque también usan alguna información extra, como eventos o movimiento. El segundo grupo se basa en detectar a los usuarios en silla de ruedas usando características discriminativas, normalmente color e

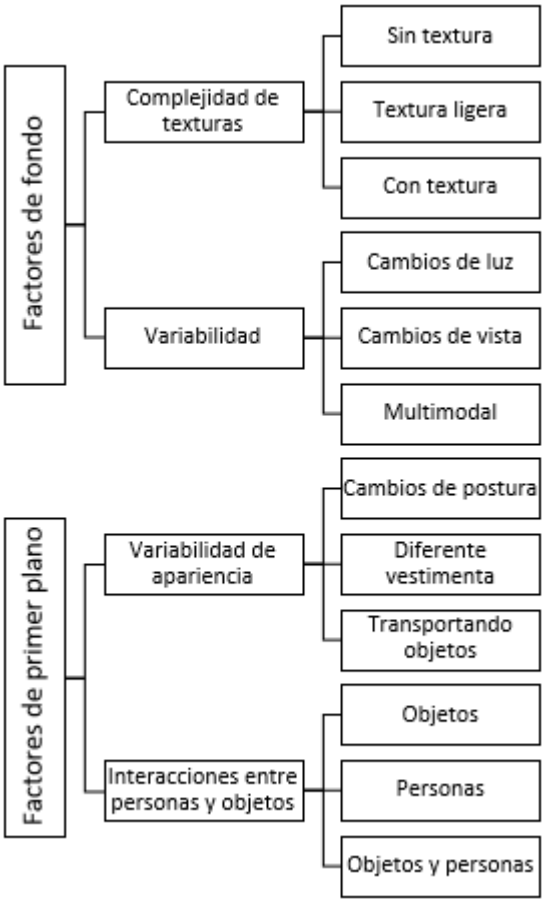



Figura 2.2: Factores críticos que afectan al detector. Fuente:[2]

Histograma de Gradientes Orientados (HOG, por sus siglas en ingles). El primer modelo que se menciona es [9], el modelo detectado y rastreado en este algoritmo consiste en dos ruedas paralelas en contacto con el suelo y una cabeza en el eje perpendicular a la recta que une ambas ruedas. El método de la detección de sillas de ruedas consiste en:

1. Sustracción de fondo: elimina la información de fondo, dejando solo los objetos detectados.

2. Skin detection: detecta la piel, con el fin de encontrar la cara.

3. Transformada de Hough: primero se utiliza un detector de Canny para extraer los bordes, y después se extraen las elipses gracias a una variación eficiente de la transformada de Hough

Otro modelo de este grupo es [10], en el que detecta las ruedas, y aprovecha la forma geométrica en 3D de la silla de ruedas, es decir, determina la localización y la orientación de la silla.

El segundo grupo, sin embargo, tiene como objetivo encontrar características discriminativas para detectas a personas en silla de ruedas. Un ejemplo de este grupo es [11], que después de una sustracción de fondo, se aprovecha de la capacidad de poder detectar por partes en orden vertical (piernas, torso, cabeza), y por último usa visión estéreo. Con esto hace que el modelo sea más resistente a diferentes sillas de ruedas.

Cabe destacar otro método en este grupo como es [12], que considera dos descriptores, HOG e Histograma de Contraste de Contexto (CCH), en el que se divide cada región en distintos cuadrantes y se mide el contraste. Este método usa un detector con CDT, en el que la entrada se divide la imagen en regiones, construyendo una pirámide gaussiana, disminuyendo la resolución en cada nivel. Luego, una ventana deslizante escanea la imagen original y cada uno de los niveles. También utiliza otro método, como es la detección con información histórica de seguimiento, que usa la relación entre frames adyacentes para incrementar la eficiencia de la detección de silla de ruedas.

# Capítulo 3

# Desarrollo

En este capítulo se tratará el proceso a seguir para la realización de este trabajo. La arquitectura llevada a cabo es la misma que la del detector de silla de ruedas explicado en el capítulo 2. Por ello, en primer lugar, se empezará explicando el detector de personas que se ha empleado y por último el modelo de persona sentada y el modelo de persona en silla de ruedas. Tanto para el detector de usuarios en silla de ruedas como para el detector de personas sentadas, el proceso de detección es el mismo, para ambas se compara toda la imagen con los modelos creados, y asigna una puntuación a cada posición de que ahí haya una persona, si el score obtenido es más alto que el umbral, entonces lo clasificará como persona. En nuestro caso el detector empleado es DTDP, descrito en [13].

## 3.1. Sistema de detección

El sistema elegido para llevar a cabo este trabajo, ha sido DTDP ("Detection Discriminatively Trained Part-Based Models"). En la detección de objetos, el principal problema se debe a la alta variabilidad de los objetos. Estas variaciones no solo se deben al punto de vista o la iluminación, sino que engloba otras variaciones debidas a deformaciones y a la variabilidad intraclase y otras propiedades visuales. Esto genera la necesidad de que el sistema de detección de objetos represente una alta variabilidad. Con el fin de subsanar este problema, el algoritmo elegido plantea que dado un objeto, este se puede dividir en partes, teniendo cada una de ellas ciertas propiedades comunes a otras regiones, y añade una componente deformable que puede ser caracterizada por la conexión entre pares de partes cercanas.

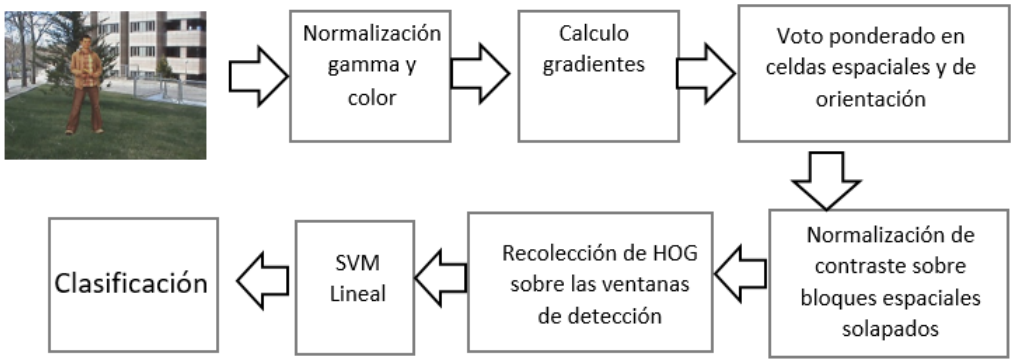



Figura 3.1: Proceso del algoritmo HOG. Fuente: [14]

A continuación se realizará una breve explicación de los algoritmos que componen la base de DTDP, como son los algoritmos HOG, Latent SVM, y por último se describirá DTDP

### 3.1.1. Histogram of Oriented Gradient (HOG)

Este método consiste en la evaluación de histogramas locales normalizados de las orientaciones de los gradientes de una imagen. Para ello, en [14] se propone el siguiente procedimiento.

La idea principal es que la apariencia de un objeto y la forma pueden caracterizarse considerablemente bien por la distribución de los gradientes de intensidad locales o por la dirección de sus bordes. En la práctica, como se ve en a figura 3.1, esto se implementa dividiendo la imagen en pequeñas regiones (celdas), y calculando para cada una su HOG. Para mejorar su invariancia a la luminosidad, sombras, etc., se realiza inicialmente una normalización de contraste. Esto se puede hacer mediante la acumulación de una medición de la "energía" de los histogramas locales de regiones más grandes, denominadas "bloques", y usan los resultados para normalizar todas las celdas del mismo bloque. Los bloques de descriptores normalizados son los descriptores de los Histogramas de Gradientes Orientados.

Esta representación captura los bordes o la estructura de gradientes, lo que es muy característico de formas locales, y con un grado de invariancia fácilmente controlable.

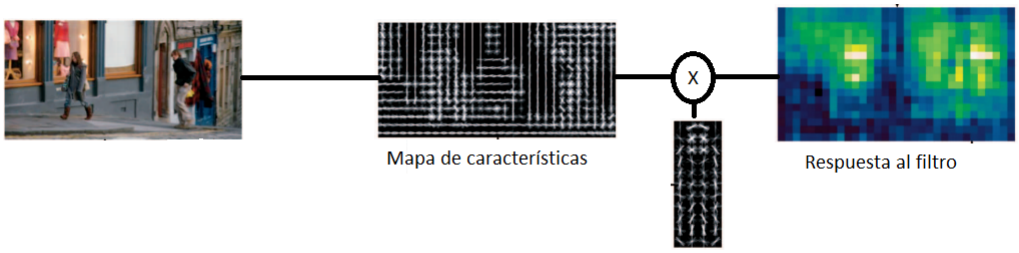



Figura 3.2: Proceso de detección por HOG. Fuente: [13]

Gráficamente, para la detección, sigue el proceso que se muestra en la figura 3.2, en el que sobre el HOG de la imagen completa, lo puntúa respecto al modelo obtenido, dando regiones en las que tenga un gran parecido.

### 3.1.2. Latent Support Vector Machine (Latent SVM)

Tal y como se menciona en [13], se considera un clasificador que puntúa una muestra $x$ con una función con la siguiente forma:

$$f_\beta(x) = \max_{z \in Z(x)} \beta \cdot \Phi(x, z)$$

Donde $\beta$ es un vector de parámetros de modelo y $z$ son valores latentes. El conjunto $Z(x)$ define los posibles valores latentes para una muestra $x$. Una etiqueta binaria para $x$ se puede obtener mediante una umbralización de una puntuación.

En relación con los clásicos SVM, se entrena $\beta$ a partir de muestras etiquetadas $D = ((\langle x_1, y_1 \rangle, ..., \langle x_1, y_1 \rangle)$ donde $y_i \in \{-1, 1\}$, reduciendo así al mínimo la función buscada a

$$L_D(\beta) = \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^{n} \max(0, 1 - y_i f_\beta(x_i))$$

donde $\max(0, 1 - y_i f_\beta(x_i))$ es la pérdida articulada estándar, y la constante $C$ controla el peso relativo del término de regularización.

Se debe tener en cuenta que si tan sólo hay un único posible valor latente para cada muestra ($|Z(x_i)| = 1$), entonces $f_\beta$ es lineal en $\beta$ y se obtiene el SVM lineal como un caso especial de SVM latente.

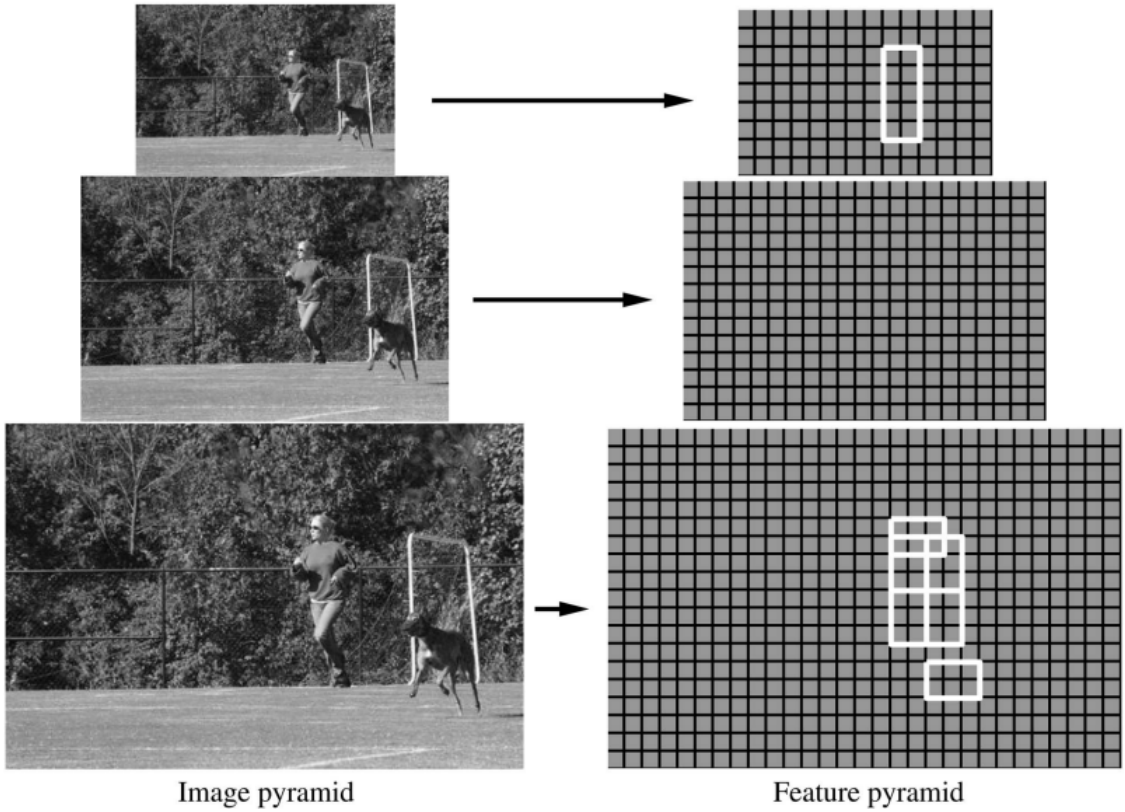



Figura 3.3: Pirámide de imágenes y características. Fuente: [13]

### 3.1.3. Discriminatively Trained Part-Based Models

En este apartado se describirá el detector DTDP, en él, como se expone en [13], los filtros se aplican sobre un mapa de densidad de características. Lo que se busca es definir una puntuación en las diferentes posiciones y escalas en una imagen. Esta puntuación se realiza usando una pirámide de características que especifica, para un número de escalas determinado, los mapas de características de cada escala, como se ve en la figura 3.3.

A diferencia del sistema explicado en [14], que usa un solo filtro para definir el modelo de un objeto, este sistema detecta objetos calculando la puntuación de cada posición y de cada escala de la pirámide de características HOG para cada parte del modelo y umbralizando la puntuación.

En la figura 3.4 se puede ver el modelo que se va a emplear. Este modelo consta de

tres filtros distintos, el filtro root ("a") que es equivalente al modelo obtenido en 3.1.1. El segundo filtro ("b"), es un filtro por partes a mayor resolución, y el tercero ("c"), representa el "coste" que supone una desviación de la posición real con respecto la posición ideal que representa el modelo. Se puede observar cómo a medida que se aleja de la posición lógica, se va "aclarando" el filtro, lo que supone un coste mayor.

También en la figura 3.4 se muestra el proceso de un detector DTDP. Inicialmente, calcula el mapa de características de la imagen a múltiples resoluciones. A continuación para resolución original, calcula la puntuación del mapa de características para el filtro root, dando como resultado la respuesta al filtro, siendo las zonas más claras las que mayor puntuación tienen. También para el mapa de características al doble de resolución, se calcula su puntuación para cada parte del filtro. Por último, la respuesta a esos filtros se suman teniendo como resultado la combinación de todas las puntuaciones obtenidas. Esta puntuación final responde a la ecuación:

$$\sum_{i=0}^{n} F' \cdot \Phi(H, p_i) - \sum_{i=0}^{n} d_i \cdot \Phi(dx_i, dy_i) + b_i$$

Donde $F'$ es el vector obtenido por la concatenación de los vectores de ponderación, $\Phi(H, p_i)$ es un vector producto de la concatenación de los vectores de características de una subventana, $d_i$ es el desplazamiento respecto su posición de anclaje y $\Phi(dx_i, dy_i)$ son las características de deformación.

## 3.2. Desarrollo de modelo de persona sentada

En este apartado se desarrollará el proceso seguido para realizar el modelo de personas sentadas. Para ello se describirá cada paso seguido.

Como ya se ha explicado en el capítulo 2, la realización de un modelo fiable es uno de los factores más críticos en la detección de personas. Por ello se ha tenido que realizar un dataset, que contase con el suficiente número de fotos como para que el resultado fuese un modelo robusto. También se cuidó tener una base de datos muy dispar, ya que, como se menciona en 3.1, el modelo debe tener cierta aceptación a la variabilidad, por lo que el conjunto debe albergar gran variedad de posturas y diferentes puntos de vista, así como información de segundo plano heterogénea. Un conjunto de imágenes poco variado, puede inducir a un modelo poco flexible a cambios.

En total se recopilaron un total de 531 imágenes, apareciendo en cada una más de

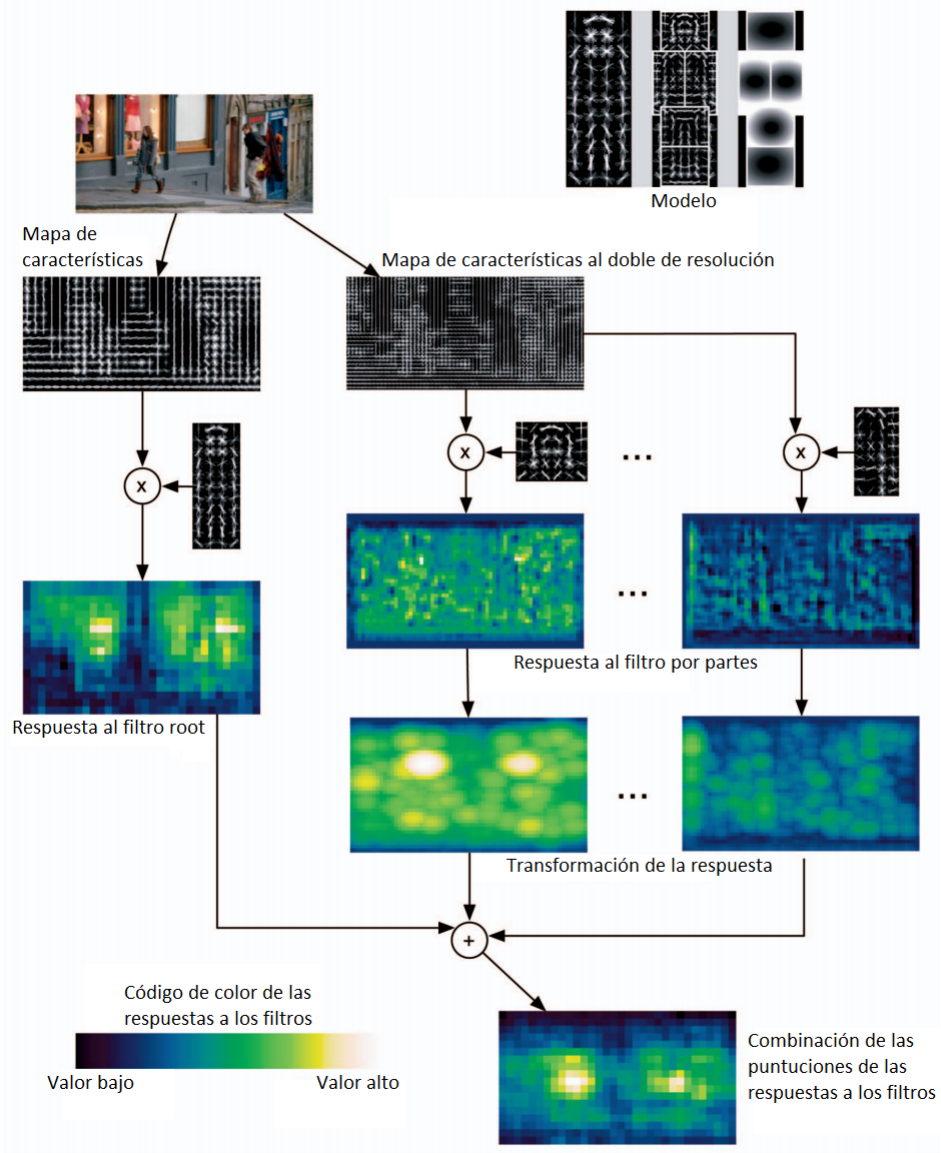



Figura 3.4: Proceso detector DTDP. Fuente:[13]

Figura 3.5: Ejemplo visual de una imagen utilizada

una persona, lo que acaba resultando un mayor número de usuarios para construir el modelo. Una vez obtenidas estas imágenes, en este caso 531, se procedió a anotarlas. Estas anotaciones contienen las características principales de la imagen. Estas características son: nombre del archivo, tamaño de la imagen (ancho, alto y profundidad), e información de cada objeto, en nuestro caso, de cada persona que aparezca en la imagen. De estos objetos la información que se aporta es el nombre, que, al haber anotado solo las personas sentadas, es el mismo para todos, siendo este *sittinguser*, si está ocluido, es decir si está presente en la imagen algún objeto que no permite ver al usuario entero, si está truncado, que significa si el usuario sentado no aparece entero debido a que rebasa los límites la imagen, y por último bounding box, que se corresponde con el tamaño del objeto y la posición, para ello se aportan las coordenadas de la columna inicial, de la fila inicial, de la columna final y de la fila final. La figura 3.5 se corresponde con una imagen del dataset recopilado. En ella se puede observar que aparecen tres usuarios interesantes. También se puede apreciar que un objeto se encuentra entre el punto de vista de la imagen y la persona del centro, provocando que las piernas estén ocultas, por lo que se anotará como ocluido.

Una vez anotadas todas las imágenes, se procedió a realizar el modelo. Como el sistema utilizado es el sistema DTDP, el modelo se compone de tres filtros, un fitro root, realizado mediante HOG, un filtro por partes, y un filtro que represente la penalización por desviación.

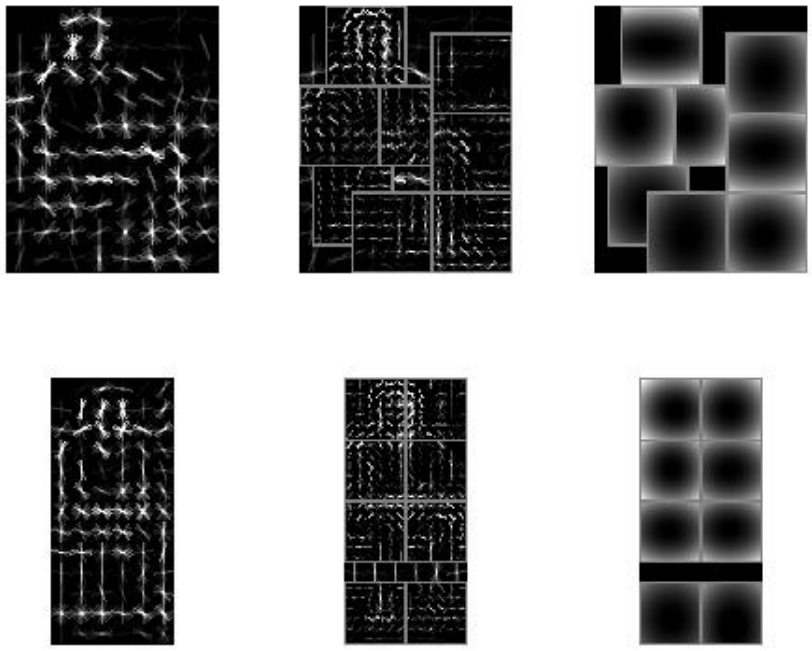


Figura 3.6: Modelo de persona sentada

La figura 3.6 muestra el modelo resultante. Se trata de un modelo múltiple en el que aparecen seis filtros. Los filtros de la parte superior corresponden, como se puede ver, al perfil lateral de la persona sentada, que incluye el filtro root, el filtro basado en partes, y el coste de la desviación. Los filtros de la parte inferior, sin embargo, competen con el perfil frontal de la persona. Es interesante observar como en el filtro por partes se observa un mayor grado de detalle que en el filtro root.

## 3.3. Desarrollo de modelo de personas en silla de ruedas a partir de imágenes sintéticas

Resulta interesante estudiar la viabilidad de realizar un detector con imágenes sintéticas. En este apartado se combinarán imágenes de personas erguidas con imágenes de sillas de ruedas con diferentes métodos que se detallarán más adelante. La posibilidad de obtener un detector fiable de esta forma permitiría evitar el coste de tener que grabar un gran número de imágenes, ya que, combinándolas resultaría: $\#Imágenes = \#Imágenes_{silla} \cdot \#Imágenes_{persona}$. En este caso, en lugar de tener que grabar 3600 imágenes (como se necesitó para realizar el modelo de silla de ruedas), tan solo serían necesarias aproximadamente 90 fotos de persona y 40 de silla de ruedas y, de forma que combinando las imágenes entre ellas se obtendría un total de 3600 imágenes.

Con el fin de obtener unos resultados de detección equiparables, se decidió alcanzar una cantidad de imágenes parecida al usado para generar el modelo de persona en silla de ruedas original. Para aproximarse a este número se consiguieron 75 imágenes de persona, y 45 de silla de ruedas, resultando un total de 3375 imágenes, que es un número muy cercano al usado para el modelo de persona en silla de ruedas original.

Para la realización de este trabajo, el laboratorio de investigación *Video Processing and Understanding Lab* facilitó diez ficheros de vídeo, en los que se había segmentado el frente (personas) del fondo. De estos vídeos se extrajeron 75 fotos de personas erguidas. Esta selección se realizó con la intención de que cumpliesen ciertas características de manera que al combinarlas, fueran lo más parecido posible a una imagen real. Estas características son:

- El frame seleccionado no podía estar ocluido.
- Las piernas no debían estar muy separadas.

- La persona debía de estar de cara, o como mucho, de perfil.

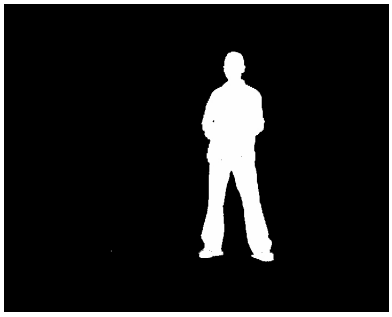
Aparte de esta selección, se reunieron 45 fotos de silla de ruedas de diversas fuentes. Una vez obtenidas, se seleccionaron parches tanto del torso como de las piernas y se combinaron con las imágenes de las sillas de ruedas con tres métodos diferentes, creando así tres dataset. Estos métodos son:

1. Combinación básica: se seleccionan tanto el torso como las piernas de la persona y se sitúan donde intuitivamente debieran estar. Se tuvo especial cuidado con que las imágenes a combinar cumplieran unas proporciones lo más realistas posible. Para ello, se debió anotar cada imagen de silla de ruedas, y cada imagen de persona, aunque en este caso solo se anotó la el torso y la parte inferior de las piernas. El proceso de anotación es el seguido en la sección 3.2. Aparte de esta anotación, también se anotó por separado las medidas tanto de la cadera como la distancia de una rodilla a otra, para ambos conjuntos (silla de ruedas y persona), y a partir de estas anotaciones se calculó el factor de redimensionamiento de la imagen. Un ejemplo de la imagen obtenida es la figura 3.8a.

2. Combinación con suavizado de bordes: como se ha explicado en la sección 3.2, el modelo se calcula mediante HOG, es decir, realiza el modelo en base a los bordes del objeto. En la figura 3.8a se puede observar que debido a la combinación de imagen, los bordes de los parches del cuerpo y las piernas son muy significativos, circunstancia que se desea evitar, ya que no se corresponde con la realidad. Para tratar de disminuir el efecto que esos bordes tan marcados producen sobre el modelo, se decidió suavizar los bordes. Este suavizado se realizó mediante un filtro paso bajo de media, de nueve filas por nueve columnas. El resultado de este suavizado se puede puede apreciar en la figura 3.8b.

3. Combinación con aplicación de máscaras: para dotar de más realismo al conjunto de imágenes, se trató de eliminar el fondo correspondiente a la imagen de persona erguida, haciendo que los parches de la persona se ajusten a su silueta.de tal forma que la imagen final fuera el resultado de combinar la imagen de silla de ruedas con el cuerpo de la persona y las piernas, como se puede ver en la figura 3.8c. Para eliminar la información de fondo se hizo uso de las máscaras aportadas. Un ejemplo de estas máscaras empleadas es la figura 3.7. Esas máscaras responden a la silueta de la persona, teniendo el interior de la silueta valor uno, y el exterior

(a) Ejemplo visual de una imagen empleada



(b) Ejemplo visual de una máscara empleada para eliminar la información de fondo

Figura 3.7: Ejemplo visual de una imagen y su respectiva máscara, extraídas de las secuencias cedidas por el laboratorio de investigación *Video Processing and Understanding Lab*

valor cero. Aprovechando estas máscaras se juntaron ambas imágenes mediante: $I_{final} = I_{silla} \cdot (1 - I_{mask}) + I_{persona} \cdot I_{mask}$ .

Después de completar los tres dataset, se crearon los modelos usando el código de entrenamiento del detector DTDP, dando lugar a los resultados de las figuras 3.9, 3.10 y 3.11.

Observando los tres modelos finales se puede apreciar el parecido entre ellos. A pesar de las diferentes técnicas, el resultado final es muy similar, aunque tienen pequeñas diferencias como por ejemplo, se aprecia mejor la región de la cabeza con la técnica de combinación por máscara. También se puede ver en el filtro por partes, la diferencia entre la combinación por suavizado y la básica, ya que en la primera de ellas, aparecen más gradientes a causa de los bordes. Otra diferencia entre la combinación con aplicación de máscara y las otras dos, es el mayor grado de detalle que se aprecia en la silla. Resulta interesante ver cómo el borde inferior de la parte superior del cuerpo, y el borde superior de las piernas son detectados como borde e incluidos en el modelo. Además, se puede apreciar el parecido con el modelo entrenado con imágenes reales, mostrado en la figura 3.12.

(a) Ejemplo de imagen por combinación básica (b) Ejemplo de imagen con suavizado de bordes (c) Ejemplo de imagen aplicando máscara

Figura 3.8: Ejemplo de imagen de los Datasets creados

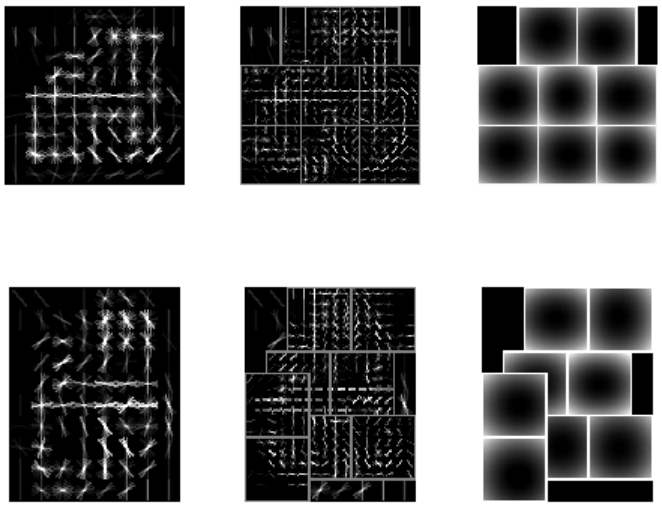


Figura 3.9: Modelo obtenido por entrenamiento a partir de imágenes sintéticas por combinación básica

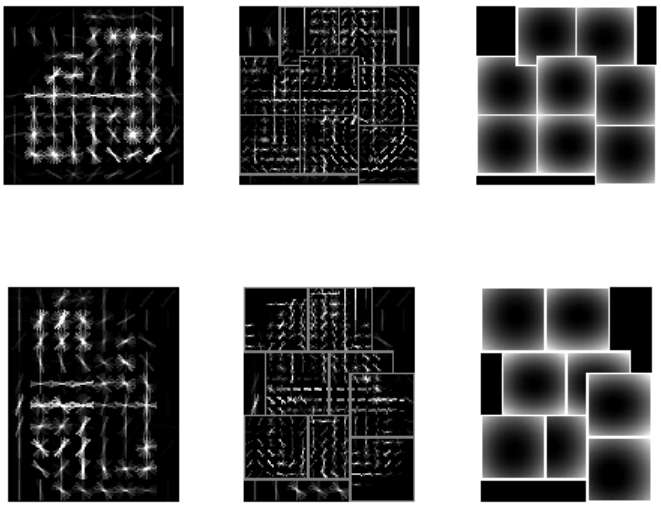


Figura 3.10: Modelo obtenido por entrenamiento a partir de imágenes sintéticas por combinación con suavizado de bordes

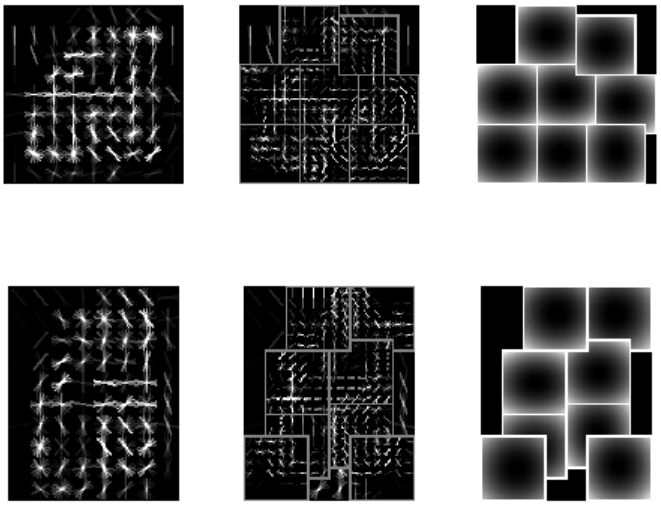


Figura 3.11: Modelo obtenido por entrenamiento a partir de imágenes sintéticas por combinación con aplicación de máscaras

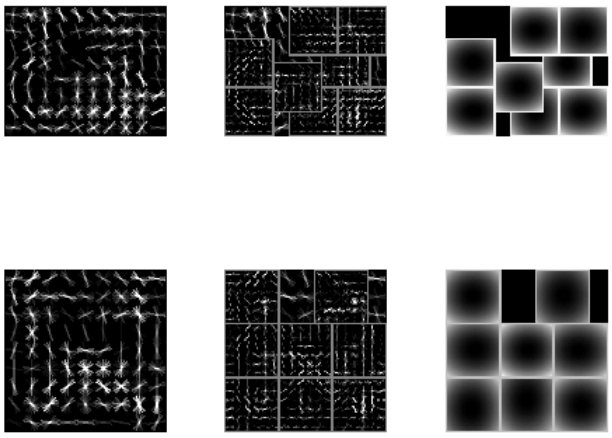


Figura 3.12: Modelo obtenido por entrenamiento a partir de imágenes reales

# Capítulo 4

# Experimentos

Este capítulo trata de abordar los experimentos realizados. Para entender mejor todo el proceso y los diferentes resultados, se hará una breve descripción del marco de evaluación en el que se basa este trabajo, donde se detallará los dataset utilizados y las métricas de evaluación empleadas. A continuación, se expondrán los resultados obtenidos para el modelo de persona sentada, en el que se comparará con el modelo de persona en silla de ruedas y persona erguida, y se determinará cómo funciona el detector empleando los tres modelos. Consecutivamente se realizarán conclusiones sobre los resultados obtenidos. Por otro lado, con respecto a los modelos elaborados por imágenes sintéticas, se analizará el funcionamiento del detector para cada modelo empleado. Por último se comentará a modo de conclusión el rendimiento de los detectores.

## 4.1. Marco de evaluación

Esta sección va a dividirse en dos partes. En la primera se realizará una descripción del dataset empleado, y en la segunda se expondrán las métricas de evaluación presentes en el trabajo, utilizadas para medir el rendimiento de los detectores.

### 4.1.1. Dataset

Dado que el objetivo de este trabajo es realizar un detector en entornos hospitalarios y residenciales, el escenario elegido es una sala de una residencia de ancianos. Las siguientes grabaciones han sido facilitadas para la realización de este trabajo por el laboratorio de investigación *Video Processing and Understanding Lab.* En ellas, aparece los miembros

(a) Ejemplo de la secuencia 1

(b) Ejemplo de la secuencia 2

Figura 4.1: Ejemplo visual de las dos secuencias tratadas

laboratorio VPU tanto sentados como andando o en silla de ruedas. Se utilizaron dos vídeos grabados desde diferentes ángulos, llamados secuencia 1 y secuencia 2, ambos con una duración de 3735 frames cada uno.

Para que el análisis del rendimiento fuera válido, se tuvieron que ajustar ambos vídeos para que tuvieran la misma duración y coincidieran temporalmente, de tal forma que el detector se probase exactamente en las mismas condiciones de luminosidad y con los usuarios realizando las mismas acciones.

Es importante comentar la posición de las cámaras con el fin de comprender en la siguiente sección el resultado obtenido. Para ello se muestra un frame de cada cámara en la figura 4.1.

A partir de estos vídeos, se tuvieron que realizar unos ficheros de anotación para cada secuencia. En estas anotaciones se apuntaron los bounding boxes. Estos concretan para cada frame la posición de cada una de las personas.

Para cada secuencia se anotaron cada uno de los objetos, tanto usuario en silla de ruedas, como persona sentada o erguida. Sin embargo se tuvo que diferenciar entre estas diferentes posturas, con la finalidad de poder evaluar cada uno de los detectores disponibles (persona erguida, persona en silla de ruedas, persona sentada) en función de su competencia. Para discriminar las diferentes posturas, se separó por regiones distintas en función de la acción que se estuviera realizando, tratando cada región como un objeto distinto. En cada anotación viene la información de cada región por separado. Dicha información es:

- Nombre de la región.
- Frame inicial: primer frame en el que aparece el objeto en la postura bajo estudio.
- Frame final: frame en el que, o el objeto sale del plano visible, o pasa a realizar otra acción (por ejemplo, está sentado y se levanta).
- Descriptores de la región: en este trabajo, vacío para todos los casos.
- Información del bounding box en cada frame. Esta información es $x$ e $y$, correspondientes al vértice superior izquierdo del rectángulo que lo define, y el ancho y el alto de este rectángulo.

Otra información relevante para este trabajo, es que para objetos que quedaban ocluidos detrás de la columna, o que eran poco visibles, se decidió anotarlos siempre y cuando en la otra secuencia se visualizaran.

Estas anotaciones se realizaron utilizando el programa *Video Image Annotation Tool (VIA)*. Este programa permite recorrer frame a frame la secuencia y definir el bounding box de cada objeto de forma manual. Una vez anotada la secuencia, el programa permite exportar la información recogida. En este caso se exportaron las anotaciones en formato de texto y en XML.

### 4.1.2. Métricas de evaluación

La manera de probar el rendimiento de un detector es, dado una secuencia de prueba, y su correspondiente Ground Truth, se ejecuta el detector sobre esa secuencia y se comparan los resultados con el Ground Truth. La métrica que se va a utilizar para comparar las detecciones realizadas y el Ground Truth es precision-recall. Se trata de un método muy empleado en reconocimiento de patrones.

Como ya se ha explicado en la subsección 3.1.3, la salida del sistema utilizado son unos valores llamados score que indican la confianza de que esos objetos encontrados sean los que originalmente se buscaban. A mayor score, mayor probabilidad de que el objeto detectado sea el deseado. El método de evaluación empleado, compara los parámetros de evaluación precisión y recall para diferentes umbrales, empezando por el umbral más alto hasta el más bajo. Cada umbral representa un punto en la curva precision-recall, es decir, cada umbral tiene una precisión y un recall asociado. Para cada umbral se genera un punto de la curva.

Los valores de precision y recall son los siguientes:

$$precision = \frac{\#detecciones\,positivas\,reales}{\#detecciones\,positivas\,reales + \#detecciones\,positivas\,falsas}$$

$$recall = \frac{\#detecciones\,positivas\,reales}{\#detecciones\,positivas\,reales + \#detecciones\,negativas\,falsas}$$

La precisión es el ratio entre el número de detecciones relevantes y el número de detecciones reales mientras que recall es el ratio entre las detecciones relevantes y y el número de todos los objetos que deberían ser detectados idealmente. Por ello, el sistema generado es mejor en cuanto más se acercan ambos valores a uno.

También se medirá el área bajo la curva precision-recall, que permitirá evaluar el rendimiento de dos curvas. A mayor área bajo la curva, mejor rendimiento.

## 4.2. Resultados de persona sentada

En esta sección se van a evaluar los detectores basados en los modelos *inriaperson*, *whelchairuser* y el implementado en la sección 3.2, *sittinguser*. Para evaluar un detector se emplea el método explicado en la subsección 4.1.2, en la que se compara los objetos detectados por un detector, con el Ground Truth de la escena, donde están apuntados todos los objetos que aparecen en ella.

En este trabajo, se realizaron diversas comparaciones que se muestran en la figura 4.2, sin embargo en la siguientes subsecciones solo se estudiarán aquellas que se han considerado más interesantes. Las comparaciones que se van a estudiar son: detector de persona de pie frente Ground Truth de persona de pie y detector de persona en silla de

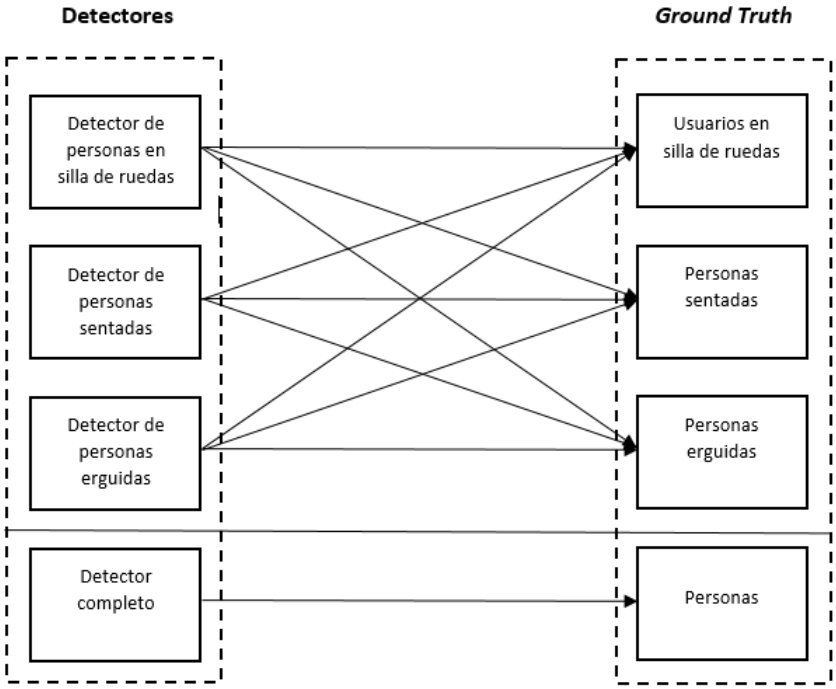



Figura 4.2: Evaluaciones realizadas

ruedas frente Ground Truth de persona en silla de ruedas, detector de persona sentada contra Ground Truth de persona sentada, detector de persona en silla de ruedas contra Ground Truth de persona sentada, y detector total frente los tres Ground Truth.

### 4.2.1. Detector de persona de pie vs Ground Truth de persona de pie y detector de silla de ruedas vs Ground Truth de silla de ruedas

En este apartado se analizará el rendimiento de los detectores implementados con los modelos *wheelchairuser* e *inriaperson*, contra el Ground Truth de usuario en silla de ruedas y persona de pie respectivamente. Es importante conocer el resultado de estos detectores de cara a comprender el funcionamiento del detector combinación de los tres detectores.

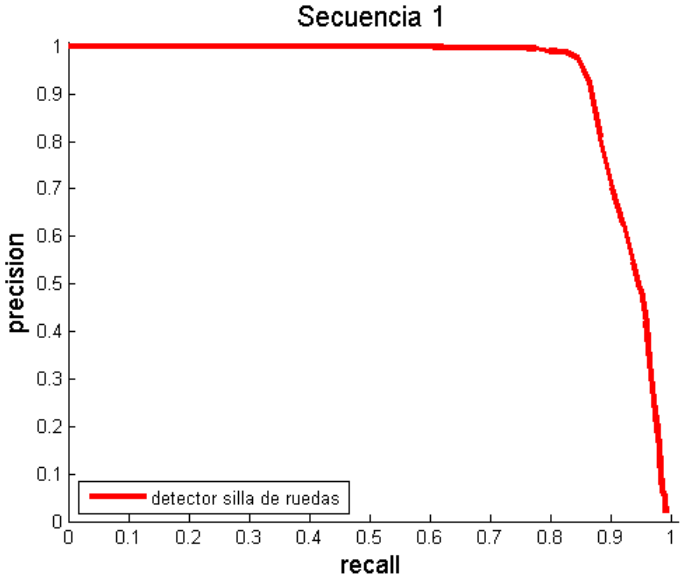
La figura 4.3a y 4.3b muestra la curva precision-recall de estos detectores respecto sus Ground Truth correspondientes. Como se puede observar, el rendimiento del detector de persona de pie en este entorno, es relativamente bajo en comparación con el detector de persona en silla de ruedas que está bastante aproximado al rendimiento ideal.

### 4.2.2. Detector de persona sentada vs Ground Truth de persona sentada

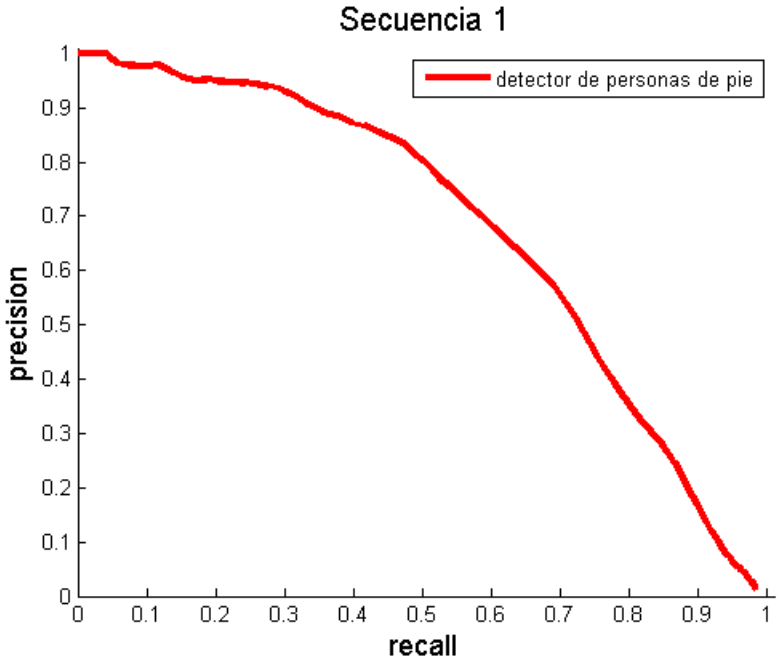
En la siguiente subsección se evaluarán los resultados del detector de persona sentada.

Como se puede ver en la figura 4.4, en la que se expone la curva precision-recall, el recall del sistema ronda el máximo, pero la precisión no supera el 60 %. A medida que disminuye el umbral, el recall aumenta y la precisión disminuye de manera progresiva. Esto quiere decir que a umbrales muy bajos sí detecta todos los usuarios sentados en la secuencia, sin embargo aun con umbrales muy altos, detecta antes otros objetos que a los usuarios sentados. Después de examinar la secuencia, se vislumbró la posibilidad de que dicho detector no discriminase unicamente a los usuarios sentados, sino que en sus detecciones también se incluyese personas en silla de ruedas. La figura 4.5, muestra un ejemplo de los objetos detectados, corroborando así la hipótesis planteada.

Por ello, se decidió evaluar el detector frente a las anotaciones de persona sentada y persona de silla de ruedas, dando como resultado la curva precision-recall de la figura 4.6. En esta nueva comparación se observa una precisión del 100 % para umbrales bajos,




(a) Curva precision-recall wheelchairuser en la secuencia 1




(b) Curva precision-recall inriaperson en la secuencia 1

Figura 4.3: Curvas precision-recall de los detectores de silla de ruedas y persona de pie contra sus propios Ground Truth

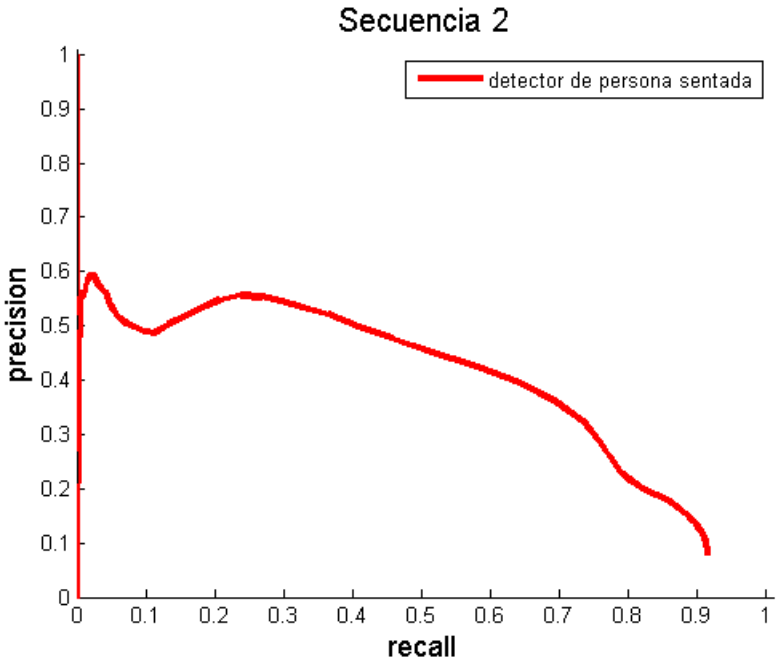



Figura 4.4: Curva precision-recall sittinguser vs. personas sentadas

Figura 4.5: Ejemplo de detección del detector de persona sentada para la secuencia 2 con un umbral de -1

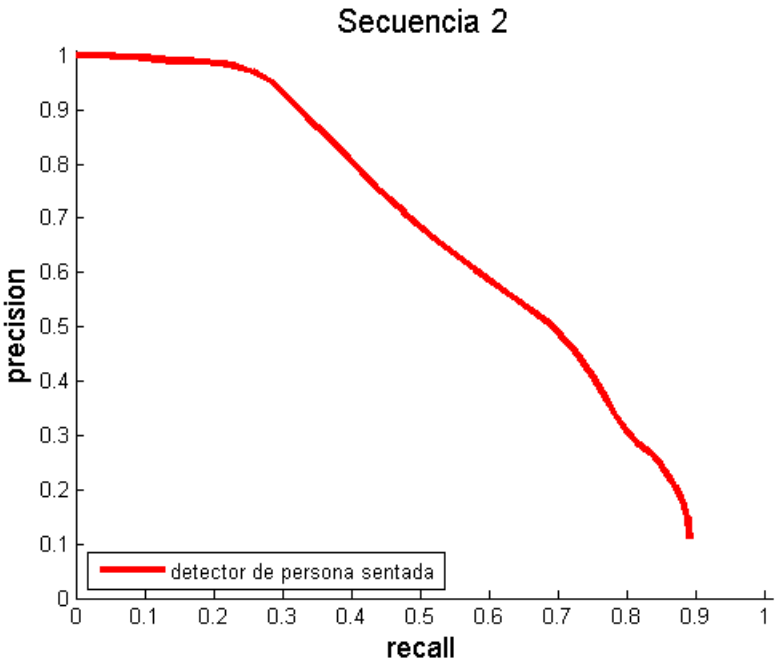



Figura 4.6: Curva precision-recall sittinguser vs. persona sentada y usuario de silla de ruedas

y un aumento considerable del área bajo la curva, lo que supone un rendimiento mucho mayor que la anterior evaluación.

### 4.2.3. Detector de persona en silla de ruedas vs Ground Truth de persona sentada

Viendo los resultados del apartado 4.2.2, donde se concluyó que el detector de persona sentada no diferencia entre usuario en silla de ruedas de usuario sentado, cabe pensar que el detector de personas en silla de ruedas y el de persona sentada sean bastante similares. Por ello se evaluó el rendimiento del detector de silla de ruedas contra el *Ground Truth* de persona sentada.

Como se aprecia en la curva precision-recall de la figura 4.7, este detector sí que es capaz de discriminar a una persona en silla de ruedas de una persona sentada.

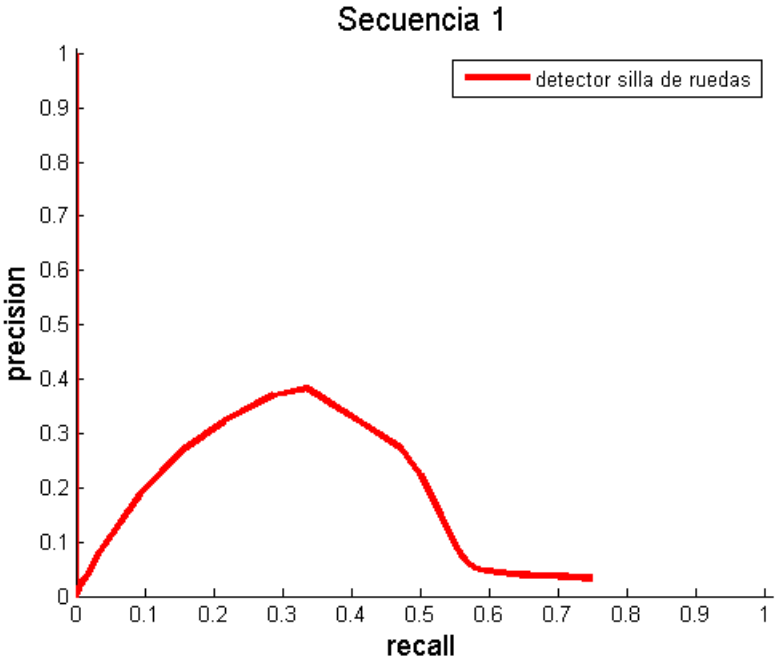



Figura 4.7: Curva precision-recall wheelchairuser vs. persona sentada

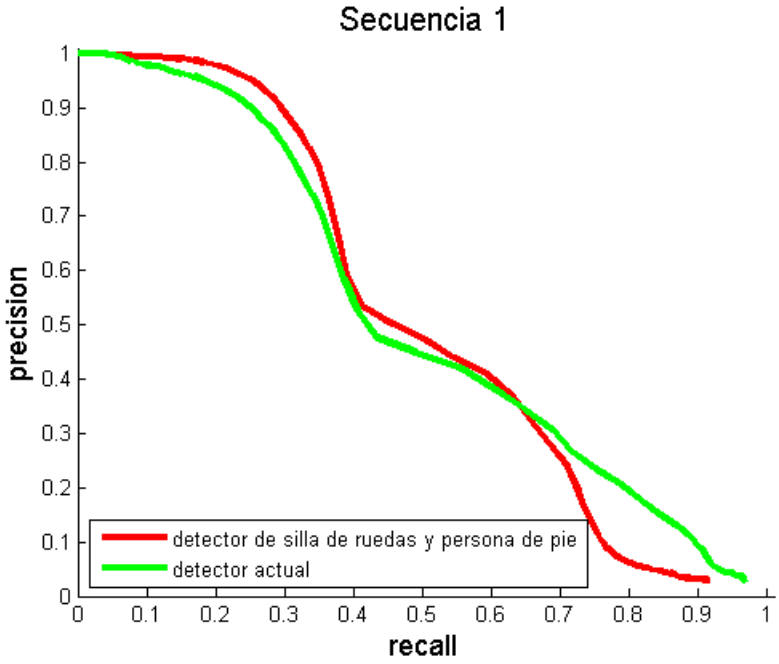
### 4.2.4. Combinacion de detectores vs Ground Truth completo

Una vez observado el rendimiento de los tres detectores, se evaluó la combinación de los tres detectores y la combinación de los detectores de persona de pie y persona en silla de ruedas frente al Ground Truth de persona sentada, de persona de pie y de silla de ruedas Esto tuvo como resultado las curvas precision-recall de la figura 4.8

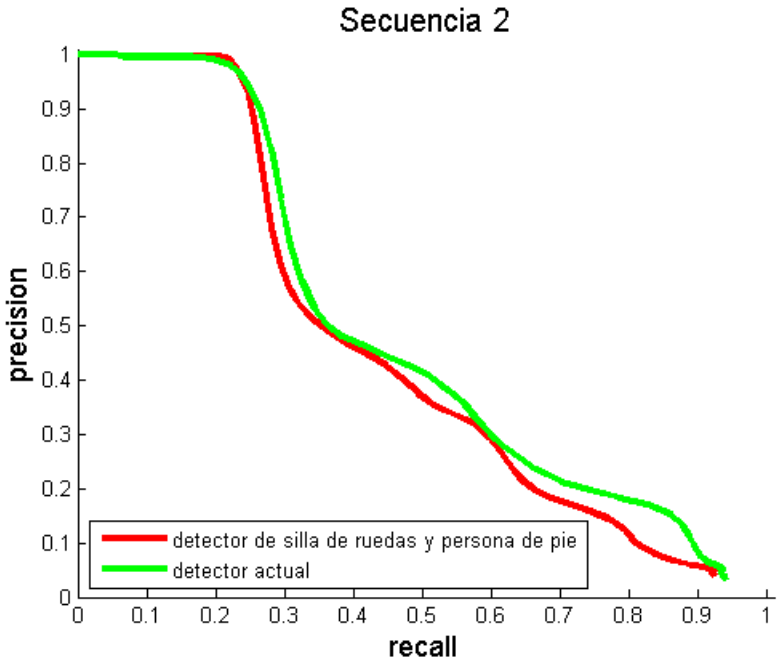
Observando los resultados se aprecia como el recall mejora levemente. Además, el área bajo la curva de los tres detectores es ligeramente mayor que el área bajo la curva de los detectores de persona de pie y persona en silla de ruedas, por lo que el detector completo tiene un mejor rendimiento.

## 4.3. Conclusiones de modelo de persona sentada

El cuadro 4.1 muestra el rendimiento de los tres detectores frente a cada Ground Truth. En concreto, en la tabla 4.1c se puede observar que el detector de personas sentadas es el que mejor rendimiento ofrece frente al Ground Truth de personas sentada, en concreto, tiene un rendimiento un 40 % mayor que los otros detectores. En vista de estos




(a) Curva precision-recall de los tres detectores vs. Ground Truth total y de los detectores de persona de pie y en silla de ruedas vs. Ground Truth total de la secuencia 1




(b) Curva precision-recall de los tres detectores vs. Ground Truth total y de los detectores de persona de pie y en silla de ruedas vs. Ground Truth total de la secuencia 2

Figura 4.8: Curva precision-recall de los tres detectores vs. Ground Truth total y de los detectores de persona de pie y en silla de ruedas vs. Ground Truth total

| Evaluación Ground Truth de persona erguida ( % del área bajo la curva) | | |
|---|---|---|
| | 1 | 2 |
| Modelo de persona erguida | 68.05 % | 70.31 % |
| Modelo de silla de ruedas | 17.84 % | 11.2 % |
| Modelo de persona sentada | 12.73 % | 07.56 % |

(a) Evaluación Ground Truth de persona erguida

| Evaluación Ground Truth de silla de ruedas ( % del área bajo la curva) | | |
|---|---|---|
| | 1 | 2 |
| Modelo de persona erguida | 22.84 % | 22.84 % |
| Modelo de silla de ruedas | 93.24 % | 76.78 % |
| Modelo de persona sentada | 47.05 % | 29.46 % |

(b) Evaluación Ground Truth de persona en silla de ruedas

| Evaluación Ground Truth de persona sentada ( % del área bajo la curva) | | |
|---|---|---|
| | 1 | 2 |
| Modelo de persona erguida | 04.18 % | 07.92 % |
| Modelo de silla de ruedas | 15.41 % | 16.49 % |
| Modelo de persona sentada | 48.03 % | 63.77 % |

(c) Evaluación Ground Truth de persona sentada

| Evaluación Ground Truth total ( % del área bajo la curva) | | |
|---|---|---|
| | 1 | 2 |
| Detector anterior | 51.79 % | 45.89 % |
| Detector con persona sentada | 52.26 % | 48.80 % |

(d) Evaluación Ground Truth total

Tabla 4.1: Rendimiento de los tres detectores frente a los distintos Ground Truth

resultados, se puede concluir que se ha mejorado el detector. Sin embargo, si se analiza el rendimiento de la combinación de detectores y el detector combinación de persona de pie y persona en silla de ruedas, se puede deducir que la mejoría es limitada, ya que el rendimiento apenas mejora un 2 %. Esto tiene su explicación en que el detector de persona sentada dista de ser fiable en este escenario, lo que puede deberse a que, como se explicó en la subsección 4.1.1, se decidió incluir en el Ground Truth a personas que aparecían ocluidas, y también a personas (sobre todo sentadas) que aparecían truncadas, de tal manera que nunca lo detecte como usuario sentado, dando un recall menor que uno.

## 4.4. Resultados del detector de persona en silla de ruedas con imágenes sintéticas

En esta sección se expondrán los resultados obtenidos por los diferentes detectores, basados en los modelos de los datasets de imágenes sintéticas por combinación básica, por combinación con suavizado de bordes y combinación con enmascaramiento frente al Ground Truth de personas en silla de ruedas.

Como se puede observar en las curvas precision-recall de la figura 4.9, para la cámara 1 el detector con mayor recall es el detector con combinación básica, y el de menor recall es el detector basado en combinación enmascarada. No obstante, la curva con mayor área, es la curva del detector de combinación con suavizado de bordes, por lo que es el que mejor rendimiento ofrece. Sin embargo, el modelo realizado por combinación con enmascaramiento es el que peor rendimiento tiene. En la otra cámara si se observan resultados más parecidos a los esperados, ya que se puede ver que el que mejor rendimiento ofrece es el detector basado en imágenes con enmascaramiento, siendo a su vez el que mayor recall alcanza.

## 4.5. Conclusión

En resumen, el recall máximo que alcanzan, no supera el 90 %, por lo que aun para umbrales muy bajos, no detecta todos los usuarios en silla de ruedas. Por otro lado, la precisión obtenida si es muy alta, esto quiere decir que para umbrales altos, no detecta otros objetos que no sean los esperados. El rendimiento ofrecido por estos detectores frente al Ground Truth de usuarios en silla de ruedas, es el que se muestra en la tabla 4.2. Como

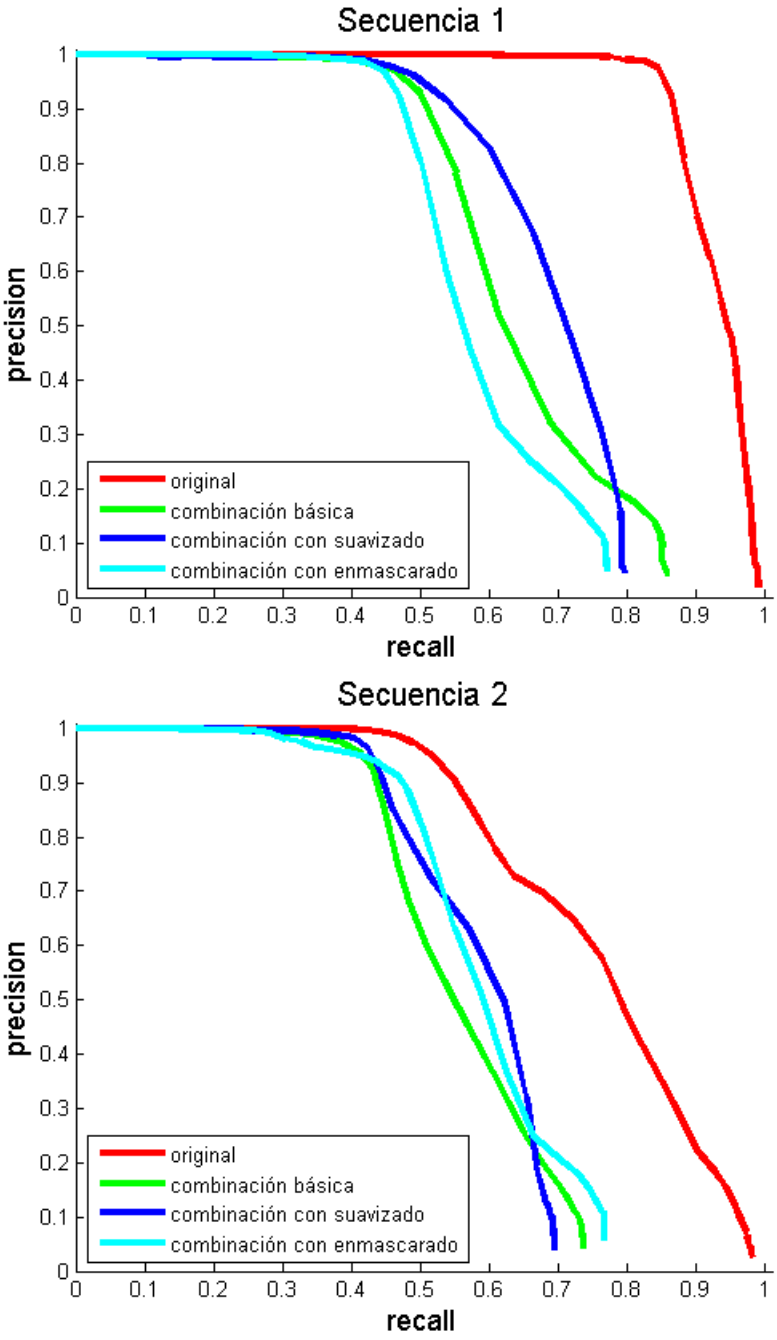



Figura 4.9: Curvas precision-recall de los tres detectores con imágenes sintéticas vs. Ground Truth de usuario en silla de ruedas

| Rendimiento frente al Ground Truth de persona en silla de ruedas ( % del área bajo la curva) | | | |
|---|---|---|---|
| | Secuencia 1 | Secuencia 2 | Average |
| Modelo de combinación básica | 64.86 % | 56.19 % | 60.53 % |
| Modelo de combinación con suavizado | 69.08 % | 58.64 % | 63.86 % |
| Modelo de combinación con máscara | 58.82 % | 59.49 % | 59.15 % |
| Modelo de imágenes reales | 93.24 % | 76.78 % | 85.01 % |

Tabla 4.2: Rendimiento de los tres detectores de imágenes sintéticas vs. Ground Truth de usuarios en silla de ruedas

se puede observar, el rendimiento para las dos cámaras es muy dispar, aunque de media el rendimiento sí es parecido. Fijándose en el rendimiento medio, el que mejor resultado ofrece es el de combinación con suavizado, mientras que el que peor es el modelo de combinación con máscara. Con estos resultados se puede concluir que el funcionamiento de los detectores es aceptable, aunque peor que el obtenido con el detector de personas en silla de ruedas original, entrenado con imágenes reales. Este resultado era de esperar, pues las imágenes que se han generado son sintéticas y son diferentes de las imágenes reales, pero pese a ello se ha obtenido un detector que es capaz de detectar de forma adecuada, con resultados de áreas bajo la curva entre el 50 % y el 60 % del total objetivo. Estos resultados se podrían mejorar generando otras imágenes sintéticas más elaboradas, pero como punto de partida es un resultado razonable.

# Capítulo 5

# Conclusiones y trabajo futuro.

## 5.1. Conclusiones.

En este trabajo se ha realizado un detector de persona sentada, que a continuación, junto con los detectores de persona de pie y silla de ruedas, se han combinado para realización de un detector que incluya estas tres posturas diferentes. Sin embargo, como se ha visto en el capítulo 4, la mejora global obtenida es muy limitada, pero la mejora para detección de personas sentadas es amplia al compararse con el resultado obtenido por el detector de personas en silla de ruedas y el detector de personas erguidas. Esta pequeña mejora global encuentra explicación en que el rendimiento del detector de persona sentada no es óptimo en este escenario. Examinando el escenario, se observa que aparecen múltiples usuarios ocluidos y truncados que el sistema no detecta.

Debido a este hecho, se puede concluir que el detector implementado obtendría mejores resultados en un escenario diferente, en la que la disposición de las cámaras no permita que existan usuarios sentados ocultos.

Otro objetivo de este trabajo era la realización de un detector de usuarios en silla de ruedas entrenado a partir de imágenes sintéticas. El resultado, como cabía esperar, es peor que el rendimiento del detector entrenado con imágenes reales, pero a cambio se ha conseguido un detector funcional sin necesidad de grabar el objeto real. Este método puede resultar útil en situaciones en las que no sea posible compilar un dataset del tipo de objeto deseado, o su obtención tenga un coste demasiado alto. Otros ejemplos en los que es aplicable esta técnica, considerando la detección de personas, podría ser para gente montando a caballo o gente en el supermercado utilizando carros de la compra.

## 5.2. Trabajo futuro.

Para mejorar el detector de persona sentada, se sugieren los siguientes cambios:

- Dado que las secuencias en las que se ha trabajado son simultáneas, se propone utilizar la información de las dos secuencias para realizar la detección, utilizando información multicámara y combinando la información obtenida por cada una de ellas.
- Otra posible mejora, es entrenar un modelo basado solo en las piernas, ya que es la parte más discriminante de una persona sentada, y es la región que en estas secuencias no aparece oculta.
- Cambio de la colocación de las cámaras.

Con la intención de aumentar el rendimiento del detector de silla de ruedas basado en imágenes sintéticas, se plantean, de acuerdo con los resultados, las siguientes propuestas:

- En vista del modelo de combinación por enmascaramiento, en el que se aprecia los gradientes obtenidos en el borde donde se une con la silla, se propone un suavizado de dicho borde.
- También se plantea la posibilidad de que, en esta combinación, incluir la zona de la cintura, con el fin de dotar de mayor realismo a la imagen resultante.
- Otros trabajos futuros con otras combinaciones como se ha comentado anteriormente: personas a caballo, personas con carros de supermercados, etc.

# Bibliografía


[1] Á. García-Martín and J. M. Martínez, "People detection in surveillance: classification and evaluation," *Computer Vision, IET*, vol. 9, no. 5, pp. 779–788, 2015. 2.1

[2] A. G. Martín, *Contributions to robust people detection in video-surveillance*. PhD thesis, Universidad Autónoma de Madrid, 2013. 2.1, 2.1.2, 2.1.3, 2.2

[3] A. Garcia-Martin and J. M. Martínez, "Robust real time moving people detection in surveillance scenarios," in *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*, pp. 241–247, IEEE, 2010. 2.1.2

[4] I. P. Alonso, D. F. Llorca, M. Á. Sotelo, L. M. Bergasa, P. R. De Toro, J. Nuevo, M. Ocaña, and M. Á. G. Garrido, "Combination of feature extraction methods for svm pedestrian detection," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 8, no. 2, pp. 292–307, 2007. 2.1.2

[5] B. Leibe, K. Schindler, and L. Van Gool, "Coupled detection and trajectory estimation for multi-object tracking," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pp. 1–8, IEEE, 2007. 2.1.2

[6] B. Leibe, A. Leonardis, and B. Schiele, "Robust object detection with interleaved categorization and segmentation," *International journal of computer vision*, vol. 77, no. 1-3, pp. 259–289, 2008. 2.1.2

[7] R. Cutler and L. S. Davis, "Robust real-time periodic motion detection, analysis, and applications," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 8, pp. 781–796, 2000. 2.1.3

[8] Á. García-Martín and J. M. Martínez, "On collaborative people detection and tracking in complex scenarios," *Image and Vision Computing*, vol. 30, no. 4, pp. 345–354, 2012. 2.1.3

[9] A. Myles, N. D. V. Lobo, and M. Shah, "Wheelchair detection in a calibrated environment," in *5th Asian Conference on Computer Vision*, pp. 2002–1, 2002. 2.2

[10] C.-A. Yang and P.-C. Chun, "Recovery of 3-d location and orientation of a wheelchair in a calibrated environment by using single perspective geometry," in *TENCON 2007-2007 IEEE Region 10 Conference*, pp. 1–4, IEEE, 2007. 2.2

[11] F. De Chaumont, B. Marhic, L. Delahoche, and C. Cauchois, "Generic method for recognition of a wheelchair, even with a low resolution-effective sensor," in *Industrial Technology, 2004. IEEE ICIT'04. 2004 IEEE International Conference on*, vol. 1, pp. 56–60, IEEE, 2004. 2.2

[12] C.-R. Huang, P.-C. Chung, K.-W. Lin, and S.-C. Tseng, "Wheelchair detection using cascaded decision tree," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 14, no. 2, pp. 292–300, 2010. 2.2

[13] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 9, pp. 1627–1645, 2010. 3, 3.2, 3.1.2, 3.3, 3.1.3, 3.4

[14] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, pp. 886–893, IEEE, 2005. 3.1, 3.1.1, 3.1.3